

W1526

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-203201

(43)Date of publication of application : 30.07.1999

(51)Int.Cl.

G06F 12/08
G06F 12/08

(21)Application number : 10-002400

(71)Applicant : HITACHI LTD

(22)Date of filing : 08.01.1998

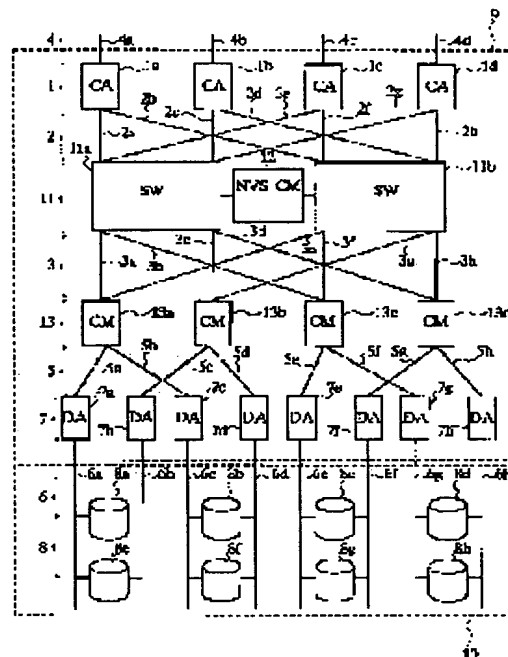
(72)Inventor : MORI KENJI

(54) ARRANGING METHOD OF CACHE MEMORY AND DATA STORAGE SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To dissolve bottleneck of an access to a cache memory caused by an increase of the number of paths or storage devices of a host device.

SOLUTION: A non-volatile cache memory (12) where write-in data from the host device are stored and a volatile cache memory 143 where data read from a disk drive 8 are temporarily stored are separately provided between plural channel I/F control circuits 1 that control plural channels I/F 4 on the side of the host device and plural disk control circuits 7 that control plural disk drives 8 in a disk drive unit 10. Moreover, the non-volatile cache memory (12) is commonly and concentratedly provided in plural data transmission paths, the volatile cache memory 13 is distributed every several data transmission paths and is arranged, and set of each capacity or throughput of the non-volatile cache memory (12) and plural volatile cache memories 13 is individually enabled.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

~1526

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-203201

(43)公開日 平成11年(1999) 7月30日

(51)Int.Cl.⁵

G 0 6 F 12/08

識別記号

3 2 0

F I

G 0 6 F 12/08

G

3 2 0

審査請求 未請求 請求項の数 3 O L (全 13 頁)

(21)出願番号

特願平10-2400

(22)出願日

平成10年(1998) 1月 8 日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目 6 番地

(72)発明者 森 健治

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74)代理人 弁理士 筒井 大和

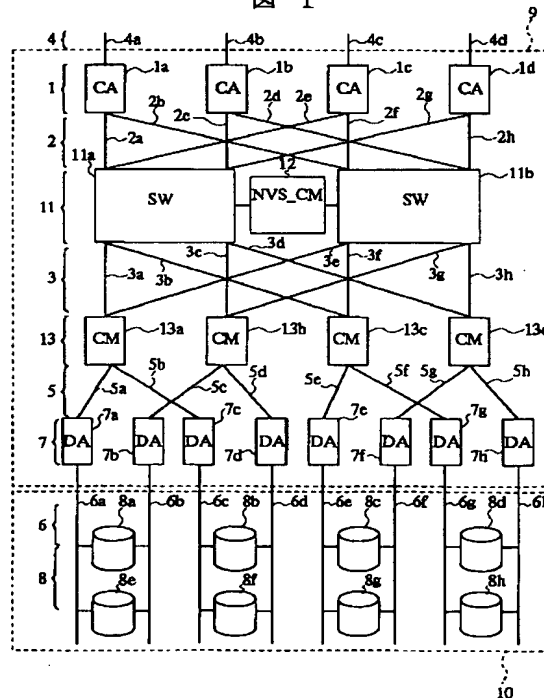
(54)【発明の名称】 キャッシュメモリの配置方法およびデータ記憶システム

(57)【要約】

【課題】 上位装置側のバス数や記憶装置数の増大に起因するキャッシュメモリへのアクセスのボトルネックを解消する。

【解決手段】 上位装置側の複数のチャンネル I / F 4 を制御する複数のチャンネル I / F 制御回路 1 と、ディスクドライブユニット 10 における複数のディスクドライブ 8 を制御する複数のディスク制御回路 7 との間に、上位装置側からの書き込みデータが格納される不揮発キャッシュメモリ 12 と、ディスクドライブ 8 から読出されたデータが一時的に格納される揮発キャッシュメモリ 13 を別個に設置するとともに、不揮発キャッシュメモリ 12 は、複数のデータ転送経路に共通に集中して設置し、揮発キャッシュメモリ 13 は、いくつかのデータ転送経路毎に分散して配置し、不揮発キャッシュメモリ 12 および複数の揮発キャッシュメモリ 13 の各々の容量やスループットを個別に設定可能にした。

図 1



【特許請求の範囲】

【請求項 1】 上位装置と、前記上位装置との間で授受される情報が格納される記憶装置との間に配置され、前記情報を一時的に保持するキャッシュメモリの配置方法であって、

前記キャッシュメモリを不揮発キャッシュメモリおよび揮発キャッシュメモリにて構成し、前記不揮発キャッシュメモリは集中的に配置し、前記揮発キャッシュメモリは分散して配置することを特徴とするキャッシュメモリの配置方法。

【請求項 2】 上位装置と、前記上位装置との間で授受される情報が格納される記憶装置との間に配置され、前記情報を一時的に保持するキャッシュメモリの配置方法であって、

前記キャッシュメモリを不揮発キャッシュメモリおよび揮発キャッシュメモリにて構成し、

前記不揮発キャッシュメモリの容量と、前記揮発キャッシュメモリの容量とが異なるように設定する第 1 の方法、

前記不揮発キャッシュメモリのスループットと、前記揮発キャッシュメモリのスループットとが異なるように設定する第 2 の方法、

の少なくとも一方の方法を用いることを特徴とするキャッシュメモリの配置方法。

【請求項 3】 上位装置との間で授受される情報が格納される記憶装置と、前記記憶装置と前記上位装置との間における前記情報の授受を制御する記憶制御装置と、前記上位装置と前記記憶装置との間に配置され、前記情報が一時的に格納されるキャッシュメモリとを含むデータ記憶システムであって、

前記キャッシュメモリは、不揮発キャッシュメモリおよび揮発キャッシュメモリからなり、

前記不揮発キャッシュメモリは集中的に配置され、前記揮発キャッシュメモリは分散して配置される第 1 の構成、

前記不揮発キャッシュメモリの容量と、前記揮発キャッシュメモリの容量とが異なる第 2 の構成、

前記不揮発キャッシュメモリのスループットと、前記揮発キャッシュメモリのスループットとが異なる第 3 の構成、

の少なくとも一つの構成を備えたことを特徴とするデータ記憶システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、キャッシュメモリの配置技術およびデータ記憶技術に関し、特に、上位装置と記憶装置との間に設けられた複数の情報転送経路にて並列的に情報の授受が行われるデータ記憶システム等に適用して有効な技術に関する。

【0002】

【従来の技術】 図 13 は、考えられる従来のディスク制御装置 109 および配下のディスクドライブユニット 110 からなるディスクサブシステムの構成の一例を示す概念図である。従来のディスク制御装置 109 では、ホスト側の入出力制御を行うチャネル制御回路 101 と、ディスクドライブ側の入出力制御を行うディスク制御回路 107 との間に、キャッシュメモリアクセスバス 105 およびキャッシュメモリアクセスバス 106 を介してアクセスされるキャッシュメモリを配置する構成が一般的である。このとき、キャッシュメモリは、電源障害に起因するデータ喪失に備えた不揮発キャッシュメモリ 102、またキャッシュメモリの障害時の補償用としての揮発キャッシュメモリ 103、を持ち単純な 2 重系として使用していた。この場合、装置の対ホストに対するバス 104 や、ディスクドライブ 108 の数が増加した場合、全データアクセスにおいて 2 つの不揮発キャッシュメモリ 102 および揮発キャッシュメモリ 103 にアクセスが発生し、キャッシュメモリにアクセス集中するため、ディスク制御装置 109 の性能におけるボトルネックとなる。

【0003】 しかし従来のディスク制御装置 109 では対ホストに対するバス 104、ディスクドライブ 108 の数が図 13 の様に 4 チャネル、8 ドライブバス程度と少なく、この様な方式である程度の性能を確保できるため、この様な方式が取られてきた。

【0004】

【発明が解決しようとする課題】 今後ディスク制御装置当たりの容量の増加、高機能化により、制御を行うディスクドライブの数が多くなり、また装置のホストに対するバス数が増加すると考えられる。

【0005】 この様な制御装置を従来の方法で構成した場合、キャッシュメモリへのアクセスが集中し、キャッシュメモリが制御装置のボトルネックとなる。このボトルネックを解消するためには、キャッシュメモリを分散配置する方法が考えられる。この時、従来の様にそれぞれ障害に備え不揮発キャッシュメモリを含む 2 重系のキャッシュメモリを用意する必要があり、この様なもので分散キャッシュを構成すると実装的に困難で、また価格的にも高価となる。

【0006】 本発明の目的は、キャッシュメモリを経由したデータ転送のスループットを向上させることが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。

【0007】 本発明の他の目的は、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。

【0008】 本発明の他の目的は、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッ

シュメモリにおける実装効率の向上を実現することが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。

【0009】

【課題を解決するための手段】本発明では、ライト時のデータを保証するための不揮発キャッシュメモリを1箇所に配置する。またリードの効率を上げるために、複数のアクセスができる様に揮発キャッシュメモリを分散配置する。この分散配置される揮発キャッシュメモリはリード時のみに使用し、データが万一電源切断等で揮発してもディスクからリードできるため、揮発性のメモリ媒体で構成可能である。

【0010】また、ディスク制御装置等の記憶制御装置ではリード／ライトの比率は対称でなく、リード4に対しライト1程度で利用される場合が多い。この性質を利用するとメモリ容量はリード／ライトの比率により不揮発キャッシュメモリはリード用の揮発キャッシュメモリほど大きくしなくても、性能に対して影響しない。またスループットも不揮発キャッシュメモリはリード用の揮発キャッシュメモリより小さくできる。しかし分散配置するリード用の揮発キャッシュメモリの容量は敏感に性能に対して影響するため、不揮発キャッシュメモリよりも容量を大きくする必要がある。この様に、キャッシュメモリの配置を分散と集中の2つの組で、それぞれの容量を要求される性能に応じた値に設定することでキャッシュメモリを最適にすることができ、またこれにより実装的、価格的にも有利に、低価格で高性能化を図ることができる。

【0011】今後、ディスク制御装置当たりの容量の増加、高機能化により、制御を行うディスクドライブの数が多くなり、また装置のホストに対するバスが増加すると考えられる。

【0012】その場合、本発明では、複数のバスに対して不揮発キャッシュメモリを1箇所に集中的に配置し、揮発キャッシュメモリを、たとえば、一つあるいは幾つかのバス毎に分散配置する。これら揮発キャッシュメモリを分散することにより、同時に対キャッシュメモリ転送が可能となり、多チャンネル、多ドライブ化への対応時に1つのキャッシュメモリに集中することなく高いスループットをもたらすことができる。また分散するキャッシュメモリは揮発性でよい実装的にも有利で、また低価格で構成可能である。さらに増設単位を分散配置される揮発キャッシュメモリのサイズを単位とすることにより、自由なスループット構成を採ることが可能となる。この様なキャッシュメモリの構成とすることで、たとえば複数バス構成のシステムにおいて低価格で高性能化を図ることが可能となる。

【0013】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【0014】図1は本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの構成の一例を示す概念図である。本実施の形態では、データ記憶サブシステムの一例として、ディスクサブシステムに適用した場合を例に採って説明する。

【0015】図1に例示される本実施の形態のディスクサブシステムは大きく、ディスク制御ユニット9、ディスクドライブユニット10、の2つのユニットより構成されている。ディスク制御ユニット9は、複数のチャンネルI/F制御回路1(1a~1d)、複数のデータバススイッチ11(11a~11b)、不揮発キャッシュメモリ12、複数の揮発キャッシュメモリ13(13a~13d)、複数のディスク制御回路7(7a~7h)、より構成され、ディスクドライブユニット10は複数のディスクドライブ8(8a~8h)より構成されている。

【0016】図1において、複数のチャンネルI/F制御回路1a~1dの各々は、複数のチャンネルI/F4(4a~4d)を介して図示しないチャンネル装置や中央処理装置(CPU)等のホスト装置に個別に接続され、チャンネルI/F4のプロトコル制御、データ変換、データ転送を行う。

【0017】複数のデータバススイッチ11は、複数のバス2(2a~2h)および複数のバス3(3a~3h)を介して接続されている複数のチャンネルI/F制御回路1と、不揮発キャッシュメモリ12および揮発キャッシュメモリ13との間におけるデータの流れの経路を切り替える役目を果たす。

【0018】複数の揮発キャッシュメモリ13は、複数のバス5(5a~5h)を介して複数のディスク制御回路7に接続され、複数のディスク制御回路7は、複数のバス6(6a~6h)を介してディスクドライブユニット10に接続されている。

【0019】不揮発キャッシュメモリ12は、ホスト側からのライト時のデータを格納するために用いる。これはライト処理の高速化の目的で、ホスト側からのライトコマンドに対して、不揮発キャッシュメモリ12への書き込みが完了した時点で、ディスクドライブ8に書き込む前に、完了を返す制御を行うため、万一、ディスクドライブ8にデータを書き込む前に停電で装置が停止した際のデータを、この不揮発キャッシュメモリ12でバックアップする目的に使用する。また、この不揮発キャッシュメモリ12に書き込む時は、可能な限り、同時に揮発キャッシュメモリ13に書き込みを行うようにデータバススイッチ11で経路の設定を行う。

【0020】揮発キャッシュメモリ13は主にリードデータのキャッシュメモリとして働き、ディスクドライブ8からデータをリードした際に、この揮発キャッシュメモリ13にデータを格納しておき、2回目以降に同じデータに対するリードに対してはディスクドライブ8から

データをリードするのではなく、揮発キャッシュメモリ13からリードすることにより高速化を図る目的に用いる。このためこの揮発キャッシュメモリ13は停電等により記憶データが揮発した場合でもディスクドライブにデータが保存されているため、揮発性のメモリ媒体で構成する事が可能である。

【0021】ディスク制御回路7は、バス6を介してディスクドライブ8に接続されており、ディスクドライブ8の制御を行うのに用いる。1つのディスク制御回路7は2つのディスクドライブ制御用のバスを有し、これにより万一の障害発生時にも交代バスを使用することによりディスクドライブ8へのアクセスが可能である。また通常時は2バスを有効に使用することで高速化の機能もはたす。

【0022】ディスクドライブ8は、データを格納するのに使用され、ディスクドライブ8へのアクセスバスは高速化、高信頼化のための2つの経路を有する。

【0023】すなわち、本実施の形態の場合には、ディスクドライブユニット10を構成する複数のディスクドライブ8は、いくつかのグループ（本実施の形態では一例として4つ）に系列化され、各系列は、複数のバス6a～6b、バス6c～6d、バス6e～6f、バス6g～6h、をそれぞれ介して、多重経路でディスク制御ユニット9に接続されている。また、複数のバス6a～6hの各々におけるデータ転送は、複数のディスク制御回路7a～7hの各々にて独立に制御される。

【0024】以下、本実施の形態のキャッシュメモリの配置方法およびディスクサブシステムの作用の一例について説明する。

【0025】図2はCPUからのリードコマンドの実行時のデータフローの一例を示した概念図である。この図2では、たとえばチャンネルI/F制御回路1aにデータリードコマンドがチャンネルI/F4aを介してCPUから入った場合について説明する。

【0026】まず、チャンネルI/F制御回路1aでCPUからのコマンドを認識する。この結果リードコマンドであることがわかる。その後、どのディスクドライブ8のデータかを判別し、判別結果に応じて、たとえば、データバススイッチ11aにおいて切り替え、チャンネルI/F制御回路1aと揮発キャッシュメモリ13aとを結合状態とする。

【0027】次にデータがキャッシュメモリ上にあるかの判定を行う。この判定はチャンネルI/F制御回路1aで行う。この結果、対象データがキャッシュメモリ上にある（キャッシュヒット）と判定した場合には、データフロー16の様に揮発キャッシュメモリ13aよりチャンネル制御回路1aにデータをデータバススイッチ11aを通じて転送することで読み出し、チャンネルI/F制御回路1aがチャンネルI/F4aを介してホスト（CPU）にデータを転送し、リードコマンドが終了する。

【0028】もしアクセス対象データが揮発キャッシュメモリ13a上に無い（キャッシュミス）場合は、データフロー17の様にディスク制御回路7a経由でディスクドライブ8aのリードを行い、データをデータバススイッチ11aを通じて転送し読み出す。この際、同時にデータを揮発キャッシュメモリ13aにライトする。これにより2回目以降のデータリードに対してはキャッシュヒットとなりアクセスの高速化を図ることができる。

【0029】図3はCPUからのライトコマンドの実行時のデータフローの一例を示した概念図である。チャンネルI/F制御回路1aにチャンネルI/F4aを介してデータライトコマンドがCPUから入った場合について説明する。

【0030】まずチャンネルI/F制御回路1aでCPUからのコマンドを認識する。この結果ライトコマンドであることがわかる。この後ライト対象がどのディスクドライブ8かを判別し、判別結果に応じてデータバススイッチ11aを切り替え、たとえば、チャンネルI/F制御回路1aと不揮発キャッシュメモリ12および揮発キャッシュメモリ13aを結合状態とする。

【0031】チャンネルI/F制御回路1aからのデータをデータフロー21の様に同時に揮発キャッシュメモリ13a、不揮発キャッシュメモリ12に転送する。これによりライトデータは不揮発キャッシュメモリ12上にライトされるとともに、分散キャッシュメモリ13aに転送される。この後にバス5aを介してディスク制御回路7aにもライトデータが転送され、さらにバス6aを介して目的のディスクドライブ8aに転送されることによって当該ディスクドライブ8aにデータが書き込まれる。

【0032】次に、停電が発生した場合の動作について説明する。チャンネルI/F制御回路1aがデータを受け取り、データフロー22の様に不揮発キャッシュメモリ12に転送が完了した時点でチャンネルI/F制御回路1aはCPUに対しコマンド完了を返す。このためCPUではデータ書き込みが完了したと判断することになる。しかしこの時点ではディスク制御ユニット9内にデータが在り、まだディスクドライブ8aに書き込まれていない状態にある。このとき停電が発生すると、転送が完了した不揮発キャッシュメモリ12と揮発キャッシュメモリ13a上の内、揮発キャッシュメモリ13a上のデータは消えてしまう。この後、電源が復旧した時に、未書き込みのデータが不揮発キャッシュメモリ12上に存在するかの確認を行い、ディスク制御ユニット9内のデータのディスクドライブ8aへの書き込みが再開される。この時、データフロー23の様に不揮発キャッシュメモリ12からのデータは、データバススイッチ11aから、バス3a、揮発キャッシュメモリ13a、バス5a、ディスク制御回路7a、バス6a、を通じて目的のディスクドライブ8aへ転送される。

【0033】図4は複数のリード/ライト発生時のデータフローの一例を示す概念図である。

【0034】これはチャンネルI/F制御回路1aにはチャンネルI/F4aを介してライトコマンドの要求が、他のチャンネルI/F制御回路1b、1c、1dにはチャンネルI/F4b、4c、4dを介してリードコマンドの要求があった場合の処理の一例を示す図である。

【0035】まず、チャンネルI/F制御回路1aの動作について説明する。チャンネルI/F制御回路1aはライトコマンドを実行するため、データバススイッチ11aを制御して不揮発キャッシュメモリ12と揮発キャッシュメモリ13aへのバスを選択する。そしてCPUからのライトデータはデータフロー34の様に、バス2a、データバススイッチ11aを経由して不揮発キャッシュメモリ12に転送され、同時に、バス2a、データバススイッチ11a、バス3a、揮発キャッシュメモリ13a、バス5aを介してディスク制御回路7aに転送され、さらにバス6aを経由してディスクドライブ8aに書き込まれる。

【0036】次にチャンネルI/F制御回路1b、1cの動作について説明する。チャンネルI/F制御回路1bでリードコマンドを実行するため、データバススイッチ11aは揮発キャッシュメモリ13bを選択する。この時、データバススイッチ11aはチャンネルI/F制御回路1aとチャンネルI/F制御回路1bの2つの要求に応じ同時に2種類のバスを形成できる。そしてリード対象のデータが揮発キャッシュメモリ13bに存在した場合（キャッシュヒット）、データフロー35の様に、揮発キャッシュメモリ13bからバス3c、バス2cを経由してチャンネルI/F制御回路1bへデータ転送が行われる。

【0037】また、チャンネルI/F制御回路1cではリードコマンドを実行するため、データバススイッチ11bは揮発キャッシュメモリ13cを選択する。そしてリード対象のデータが揮発キャッシュメモリ13cに存在した場合（キャッシュヒット）、データフロー36の様に、揮発キャッシュメモリ13cから、バス3f、バス2fを経由してチャンネルI/F制御回路1cへデータ転送が行われる。

【0038】最後にチャンネルI/F制御回路1dの動作について説明する。チャンネルI/F制御回路1dでリードコマンドを実行するため、データバススイッチ11bは揮発キャッシュメモリ13dを選択する。この揮発キャッシュメモリ13dに対象データが存在しない場合（キャッシュミス）、データフロー37の様にバス5hを経由してディスク制御回路7hにリード要求が入り、バス6hを経由してディスクドライブ8dから読出されたリードデータは揮発キャッシュメモリ13dにライトされながら、バス3h、バス2hを経由してチャンネルI/F制御回路1dに転送される。この様に揮発キャッシ

ュメモリ13を、バス毎に13a~13dのように分散配置することにより、複数のバスに対するデータ転送を同時に行うことが可能となる。

【0039】図5は複数のリード/ライトが特定のディスクドライブ（たとえばディスクドライブ8d）に集中した場合のデータフローの一例を示す概念図である。

【0040】これはチャンネルI/F制御回路1a、1bにライトコマンド、チャンネルI/F制御回路1c、1dにリードコマンドの要求があった場合の図である。

【0041】まずチャンネルI/F制御回路1aの動作について説明する。データフロー43の様に、チャンネルI/F制御回路1aはライトコマンドを実行するため、データバススイッチ11aを不揮発キャッシュメモリ12と揮発キャッシュメモリ13cへのバスを選択する。しかしこの時、リードコマンド等によりディスク制御回路7gが使用されている場合は、不揮発キャッシュメモリ12までの転送のみを先に行い、その後、ディスク制御回路7gが空いてから、停電時と同じ様に不揮発キャッシュメモリ12から、揮発キャッシュメモリ13cを経由してディスク制御回路7gにデータを転送し、ディスクドライブ8dに書き込む。図5の例の場合は、ディスク制御回路7gが空いているため同時に転送を行う。

【0042】ホストに対するコマンドの終了報告はディスクドライブ8dに書き込みが完了する前に、不揮発キャッシュメモリ12に書込んだ時点で行う。基本的には不揮発キャッシュメモリ12へは確実に書き込み、その時、ディスク制御回路7gへのバスが空いている場合は同時にディスク制御回路7gへの転送を行う処理とする。

【0043】チャンネルI/F制御回路1bの動作についても同様に、データフロー44の様に一時的に不揮発キャッシュメモリ12に書き込みを行い、その後、ディスク制御回路7hが空いてから不揮発キャッシュメモリ12から、揮発キャッシュメモリ13dを経由してディスク制御回路7hにデータを転送し、ディスクドライブ8dに書き込む。この例では不揮発キャッシュメモリ12に対する書き込みは、2つのチャンネルI/F制御回路1a、1bよりのライト要求が競合するがこれは順番に行う。

【0044】次にチャンネルI/F制御回路1cの動作について説明する。チャンネルI/F制御回路1cでリードコマンドを実行するため、データバススイッチ11bが揮発キャッシュメモリ13dを選択する。そしてリード対象のデータが揮発キャッシュメモリ13dに存在した場合（キャッシュヒット）、データフロー45の様に揮発キャッシュメモリ13dから、バス3h、バス2fを経由してチャンネルI/F制御回路1cへデータ転送が行われる。

【0045】もし揮発キャッシュメモリ13dに目的のデータが存在しない場合（キャッシュミス）は、バス6

h、ディスク制御回路7h、バス5h、揮発キャッシュメモリ13d、バス3hを経由してディスクドライブ8dよりデータを読み出す。

【0046】チャンネルI/F制御回路1dについても同様に、データフロー46の様に、まず揮発キャッシュメモリ13dにリード対象のデータがあるかの確認を行い、ある場合（キャッシュヒット）には揮発キャッシュメモリ13dより転送を行い、無い場合（キャッシュミス）はディスクドライブ8dよりデータを読み出す。この時、リードコマンドが競合するが、これは順番に行う。

【0047】従来の技術では、このような場合、単一のディスクドライブに対するリード、ライトでも分散したディスクドライブに対するアクセスでも、キャッシュメモリに対するアクセスの競合が発生し、待ちが多くなることになり、性能低下の原因となった。

【0048】しかし本実施の形態では、分散したディスクドライブに対するアクセスでは競合は発生せず高性能が得られる。また図5の様に1つのディスクドライブ8dに対するアクセスが集中するケースは、短いアクセスが多数発生するランザクション処理では少ないため、リード時の特定の揮発キャッシュメモリに対するアクセスの集中は少ないといえる。また集中管理するメモリとして不揮発キャッシュメモリ12がありライト時にアクセスが集中するが、ライトの比率が一般的にリードに比べ4分の1程度のため、それほど問題にはならない。そのため、これにより不揮発キャッシュメモリ12の容量も他の揮発キャッシュメモリ13a~13dの総和に比べて少なく済み、低価格で高性能なディスクサブシステムを実現することが可能となる。

【0049】ここでキャッシュメモリのサイズ、スループットに関する説明を行う。図5の様に不揮発キャッシュメモリ12を1つ、また揮発キャッシュメモリ13を、一例として13a、13b、13c、13dの4組み持った構成の場合を考える。ライトとリードに対する比率を1:3とする。これにより不揮発キャッシュメモリ12のサイズ、スループットはリードに用いる複数の揮発キャッシュメモリ13（13a~13d）の等価サイズ、スループットの4分の1程度ですむことになり、実装的、価格的に都合がよい。

【0050】次に複数の揮発キャッシュメモリ13a~13dから構成される揮発キャッシュメモリ13のサイズ、スループットについて考える。この時、複数の揮発キャッシュメモリ13a~13dの各々のカーバー範囲がディスクドライブバス（バス5aと5b、バス5cと5d、バス5eと5f、バス5gと5h）とした場合、このディスクドライブバスに対する集中の度合いでその値が決まる。図5の様にバス5a~5hの8組で構成され、それぞれの競合がないとする場合、等価スループットそのもので良いが、平均2つの競合が発生すると

した場合は等価スループットの2倍の性能が必要になってくる。またサイズに関しては、単純には組み数（バス5の数）で割れば良いが、データの分散の程度により値が異なる。一部に集中した場合を想定した場合にはその分多く持つことでキャッシュメモリの効果を引き出すことができる。これは確率的に値を決めることになる。

【0051】ところで、本実施の形態のように、不揮発キャッシュメモリ12と、揮発キャッシュメモリ13（13a~13d）とを分散して配置し、不揮発キャッシュメモリ12に対するデータ書き込みが完了した時点でホスト側に書き込み完了を応答する構成では、不揮発キャッシュメモリ12への書き込みデータが、必ずしも直ちに揮発キャッシュメモリ13やディスクドライブ8に反映されているとは限らない。このため、未反映の間にリード要求が発生した場合には、最新のデータが、不揮発キャッシュメモリ12、揮発キャッシュメモリ13、ディスクドライブ8のいずれに存在するかを判別する操作が必要となる。

【0052】本実施の形態では、一例として、図6に例示されるような制御情報を用いて、このような判別操作を行う。

【0053】すなわち、不揮発キャッシュメモリ12では、たとえばアクセス単位のエントリ毎に、NVS管理フラグ50（ V_N ）を設ける。本実施の形態の場合、 V_N が“0”のとき、当該エントリの書き込みデータは揮発キャッシュメモリ13に未反映であり、“1”のときは反映済である。

【0054】また、揮発キャッシュメモリ13では、たとえばアクセス単位のエントリ毎に、CM管理フラグ51（ V, A ）を設ける。本実施の形態の場合、 V が“0”のとき、当該エントリのデータに対して、不揮発キャッシュメモリ12に未反映のデータが存在し、 V が“1”のときは存在しない。また、 A が“0”のとき、当該エントリの書き込みデータはディスクドライブ8上に未反映の状態にあり、 A が“1”のときは反映済である。

【0055】なお、CM管理フラグ51においては、電源投入直後は、格納データが消失しているが、この状態では、全エントリの V および A は、ともに“0”の状態にあり、この状態では、キャッシュミスと判定され、ディスクドライブ8上からのデータリードが実行される。そして、不揮発キャッシュメモリ12に存在するディスクドライブ8に未反映のデータの当該ディスクドライブ8への書き込み操作や、ディスクドライブ8から読出されたデータの格納操作によって、 V および A は後述のように変化する。

【0056】そして、データ書き込みの際には、たとえば、図7のフローチャートに例示されるように、ホスト側（チャンネルI/F制御回路1）から到来する書き込みデータを、不揮発キャッシュメモリ12に書き込んだ

のち、NVS管理フラグ50の V_N を“0”にセットし（ステップ201）、さらにCM管理フラグ51の V を“0”にセットする（ステップ202）。その後、ホスト側にライト完了を応答する（ステップ203）。なお、ステップ202ではCM管理フラグ51の V の操作のために揮発キャッシュメモリ13へのアクセスが発生するが通常データ転送とは異なり、わずかなフラグビットの操作のみであるため、オーバーヘッドは少ない。

【0057】たとえば、上述の図5の例のように、不揮発キャッシュメモリ12以下への書き込みデータの転送は、任意契機でよく、たとえば、図8のフローチャート例示されるような手順にて行われる。

【0058】すなわち、まず、NVS管理フラグ50の V_N が“0”のエントリを不揮発キャッシュメモリ12から検索し（ステップ301）、当該データを、揮発キャッシュメモリ13に転送した後、CM管理フラグ51の V を“1”にセットする（ステップ302）。さらに、揮発キャッシュメモリ13からディスクドライブ8上に書き込みデータを転送した後、CM管理フラグ51の A を“1”にセットする（ステップ303）。最後に、NVS管理フラグ50の V_N を“1”にセットする（ステップ304）。この一連の操作は任意契機で実行可能である。

【0059】一方、任意の契機で発生するホスト側からのリード要求の処理は、一例として、図9に例示されるフローチャートのようにして行われる。

【0060】すなわち、リード要求が発生すると、まず該当する揮発キャッシュメモリ13のCM管理フラグ51がチェックされ（ステップ401）、 $A=1$ かつ $V=0$ の場合には、リード要求されたデータに対応した未反映の書き込みデータが不揮発キャッシュメモリ12に存在すると判定して、不揮発キャッシュメモリ12からデータを読み出してホストに転送する（ステップ404）。

【0061】また、ステップ401において、 $A=1$ かつ $V=0$ でないと判定された場合には、さらに、 $A=0$ かつ $V=1$ 、または、 $A=1$ かつ $V=1$ か否かを調べ（ステップ402）、この条件が成立する場合には、揮発キャッシュメモリ13のキャッシュヒットとして、揮発キャッシュメモリ13内のデータを読み出してホスト側に転送する（ステップ405）。

【0062】ステップ401、ステップ402のいずれの条件にも合致しない場合には、キャッシュミスと判定し、ディスクドライブ8からデータを読み出し、揮発キャッシュメモリ13に書き込みつつ、ホスト側にデータを転送し、CM管理フラグ51の A および V を“1”にセットする（ステップ403）。

【0063】このようなNVS管理フラグ50およびCM管理フラグ51を用いた一連の処理により、データ書き込み要求に際してのデータ書き込み動作が、不揮発キャッシュメモリ12以下の揮発キャッシュメモリ13、

さらにディスクドライブ8のどのレベルで未実行であるか否かに関係なく、ホスト側からのリード要求に対して、最新データのリードを的確に実行可能であり、たとえば、最新の書き込みデータが未反映の古いデータを誤って読出してホスト側に転送する、等の障害の発生を確実に回避することができる。

【0064】また、このような管理に際してアクセスされるデータは、高々数ビットであるため、NVS管理フラグ50およびCM管理フラグ51の操作に起因するオーバーヘッドはリード/ライト処理のスループットにはほとんど影響しない。

【0065】以上説明したように、本実施の形態のキャッシュメモリの配置方法およびデータ記憶システムによれば、複数のチャンネルI/F4やバス2を備えた多チャンネルバス化、ディスクドライブユニット10におけるディスクドライブ8の数量の増大によって多ディスクドライブ化されたディスク制御ユニット9の構成において、揮発キャッシュメモリ13をいくつかの経路毎に分散配置することで、高スループット化が可能となり、さらに、揮発キャッシュメモリ13とは別個に不揮発キャッシュメモリ12を集散的に配置して管理することで、不揮発キャッシュメモリ12および揮発キャッシュメモリ13の各々のサイズを最適に設定することができ、実装面で有利に、また低価格で高性能なディスク制御ユニット9、すなわち、ディスクサブシステムを実現することが可能となる。

【0066】なお、不揮発キャッシュメモリおよび揮発キャッシュメモリの分散配置方法としては、図1に例示された方法に限らず、たとえば、図10～図12に例示された構成を用いることもできる。なお、図10～図12において図1と共通な構成要素には共通の符号を付して説明は割愛する。

【0067】すなわち、図10の場合には、ホスト側の複数のチャンネルI/F制御回路1a～1dと、ディスクドライブ8側の複数のディスク制御回路7a～7dとが、別個に配置される不揮発キャッシュメモリ12および揮発キャッシュメモリ13を介して接続される構成としたものである。このような構成においても、上述の図1に例示される構成における効果とともに、揮発キャッシュメモリ13の制御回路をより簡略化できる、という利点がある。

【0068】図11の場合は、ホスト側の複数のチャンネルI/F制御回路1a～1dと、ディスクドライブ8側の複数のディスク制御回路7a～7fとの間をデータバススイッチ11を介して接続した構成において、ディスクドライブ8毎に系列をなす、複数のディスク制御回路7a、7b、7c、7d、7e、7f、の各系列毎に、互いに独立な不揮発キャッシュメモリ12および揮発キャッシュメモリ13の組を配置したものである。この図11の構成の場合には、ディスクドライブ8の系列毎

に、不揮発キャッシュメモリ 1 2 および揮発キャッシュメモリ 1 3 の組み合わせにおける容量やスループットの組み合わせの最適化を実現できる、という利点がある。

【0069】図 1 2 の場合には、図 1 1 におけるデータバススイッチ 1 1 を省略するとともに、ディスクドライブ 8 の各系列が、いわゆる R A I D におけるパリティグループを構成し、各パリティグループ毎に、互いに独立な不揮発キャッシュメモリ 1 2 および揮発キャッシュメモリ 1 3 の組を分散して配置したものである。この場合には、たとえば各パリティグループ毎に稼働状況が異なる場合に、当該各パリティグループ毎の不揮発キャッシュメモリ 1 2 および揮発キャッシュメモリ 1 3 の組み合わせにおける容量やスループットの組み合わせの最適化を実現できる、という利点がある。

【0070】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【0071】たとえば、データ記憶システムとしてはディスクサブシステムに限らず、記憶階層を有する一般のデータ記憶システムに広く適用することができる。

【0072】

【発明の効果】本発明のキャッシュメモリの配置方法によれば、キャッシュメモリを経由したデータ転送のスループットを向上させることができる、という効果が得られる。

【0073】また、本発明のキャッシュメモリの配置方法によれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することができる、という効果が得られる。

【0074】また、本発明のキャッシュメモリの配置方法によれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおける実装効率の向上を実現することができる、という効果が得られる。

【0075】また、本発明のデータ記憶システムによれば、キャッシュメモリを経由したデータ転送のスループットを向上させることができる、という効果が得られる。

【0076】また、本発明のデータ記憶システムによれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することができる、という効果が得られる。

【0077】また、本発明のデータ記憶システムによれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおける実装効率の向上を実現することができる、という効果が得られる。

【図面の簡単な説明】

【図 1】本発明のキャッシュメモリの配置方法が実施さ

れるデータ記憶サブシステムの構成の一例を示す概念図である。

【図 2】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるリードコマンドの実行時のデータフローの一例を示した概念図である。

【図 3】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるライトコマンドの実行時のデータフローの一例を示した概念図である。

【図 4】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおける複数のリード／ライト発生時のデータフローの一例を示す概念図である。

【図 5】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおいて、複数のリード／ライトが特定のディスクドライブに集中した場合のデータフローの一例を示す概念図である。

【図 6】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおいて用いられる制御情報の一例を示す説明図である。

【図 7】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ書き込み処理の一例を示すフローチャートである。

【図 8】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ書き込み処理の一例を示すフローチャートである。

【図 9】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ読み出し処理の一例を示すフローチャートである。

【図 10】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図 11】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図 12】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図 13】考えられる従来の、ディスク制御装置および配下のディスクドライブユニットからなるディスクサブシステムの構成の一例を示す概念図である。

【符号の説明】

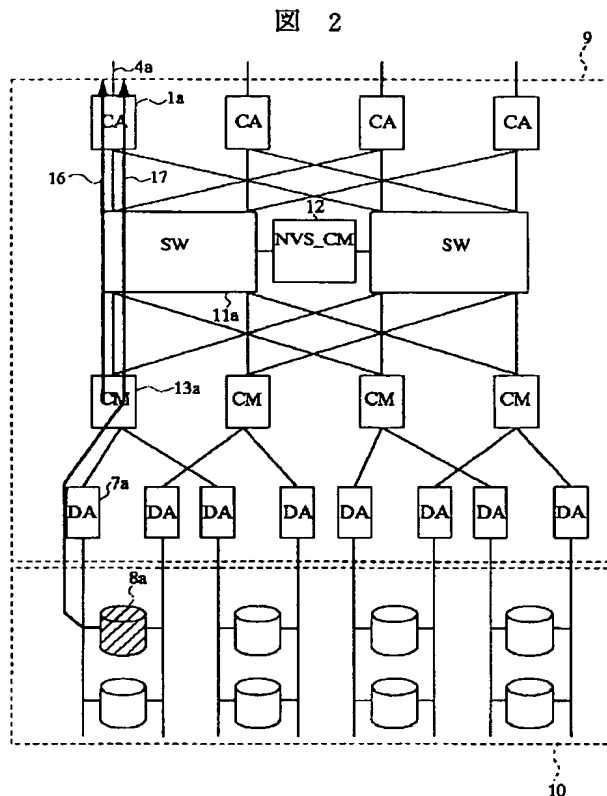
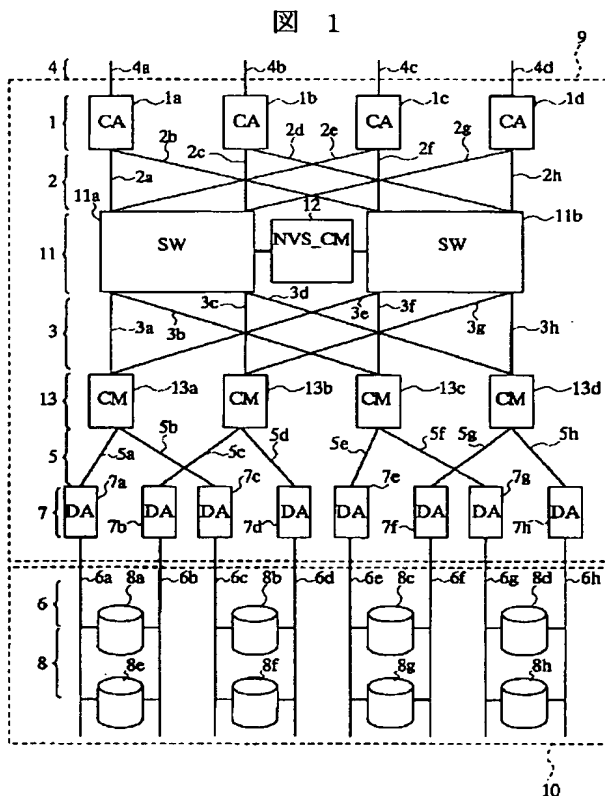
1 (1 a ~ 1 d) …チャネル I / F 制御回路、2 (2 a ~ 2 h) …バス、3 (3 a ~ 3 h) …バス、4 (4 a ~ 4 d) …チャネル I / F、5 (5 a ~ 5 h) …バス、6 (6 a ~ 6 h) …バス、7 (7 a ~ 7 h) …ディスク制御回路、8 (8 a ~ 8 h) …ディスクドライブ、9 …ディスク制御ユニット (記憶制御装置)、10 …ディスクドライブユニット (記憶装置)、11 (11 a, 11 b) …データバススイッチ、12 …不揮発キャッシュメモリ、13 (13 a ~ 13 d) …揮発キャッシュメモリ、16, 17 …データフロー、21 ~ 23 …データフ

ロー、34~37…データフロー、43~46…データ
フロー、50…NVS管理フラグ、51…CM管理フラ

グ。

【図1】

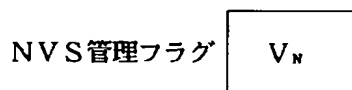
【図2】



【図6】

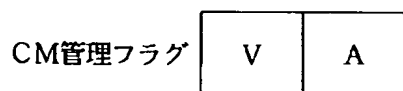
【図10】

図6



$V_N = 0$: NVS上の書き込みデータはCMに未反映。

$V_N = 1$: NVS上の書き込みデータはCMに反映済。



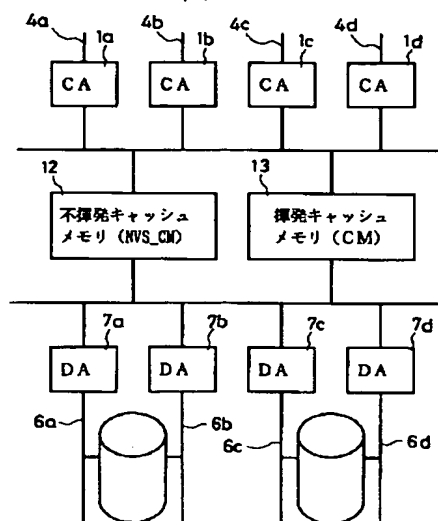
$V = 0$: NVSからCMに未反映の書き込みデータあり。

$V = 1$: NVSからCMに未反映の書き込みデータなし。

$A = 0$: CMからDISKに未反映の書き込みデータあり。

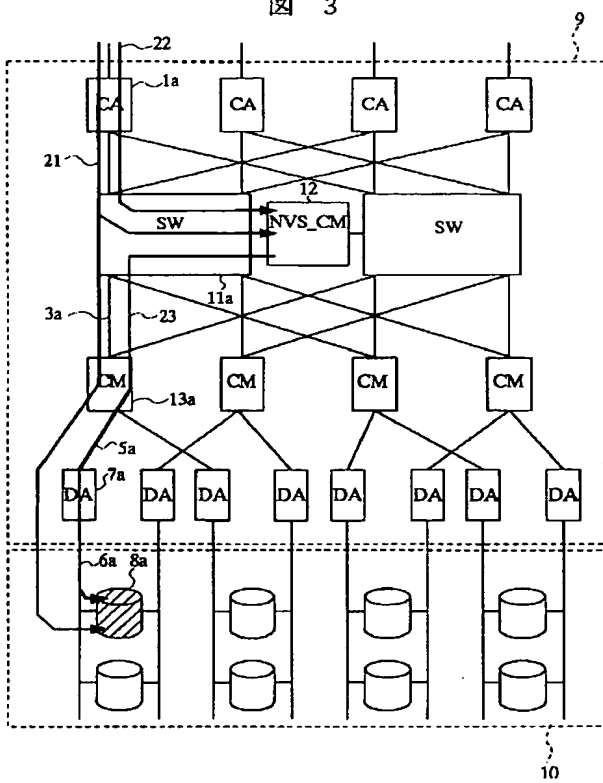
$A = 1$: CMからDISKに未反映の書き込みデータなし。

図10



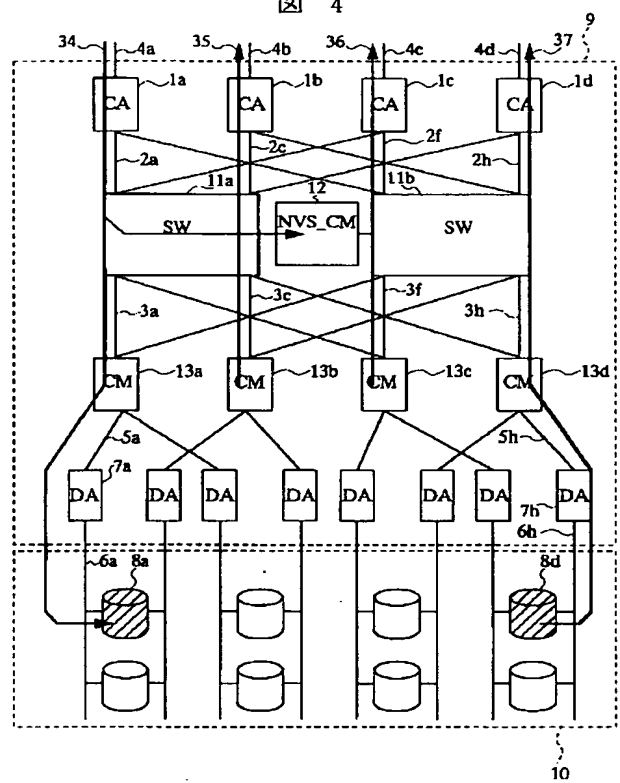
【図 3】

図 3



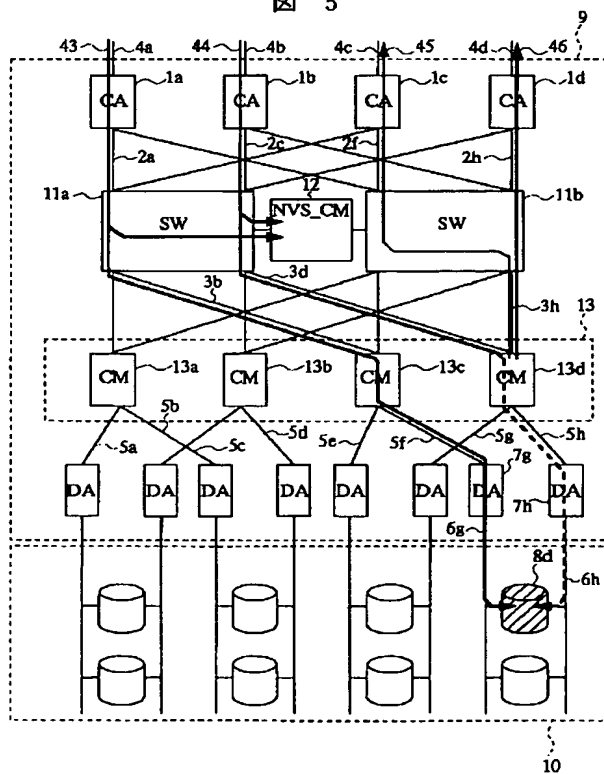
【図 4】

図 4



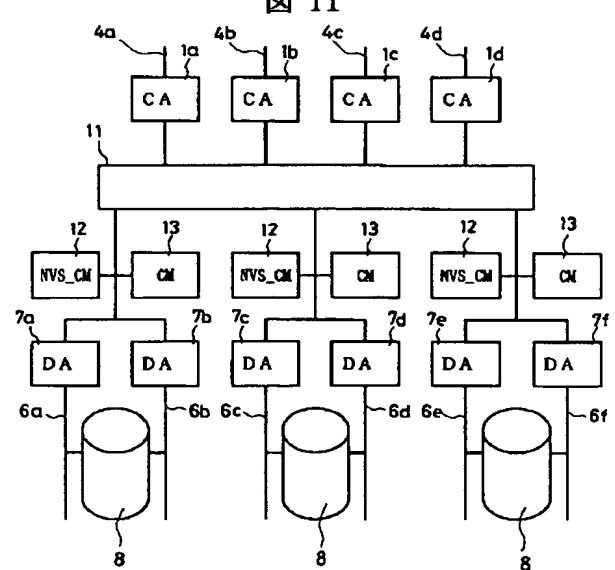
【図 5】

図 5



【図 11】

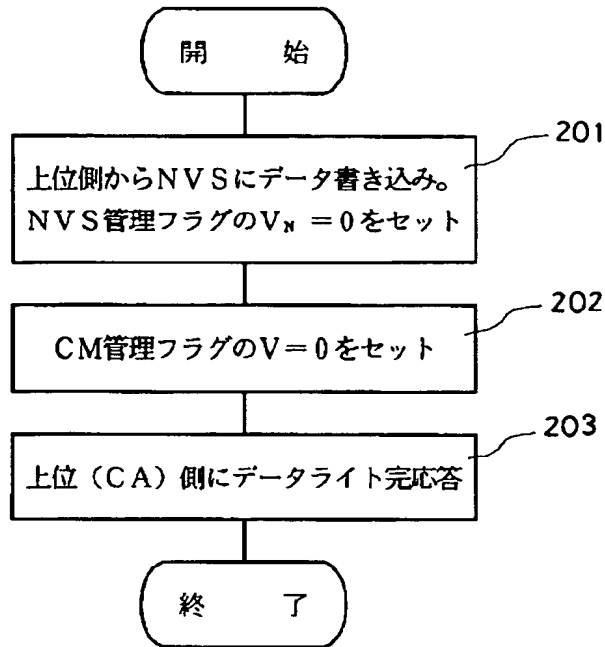
図 11



【図 7】

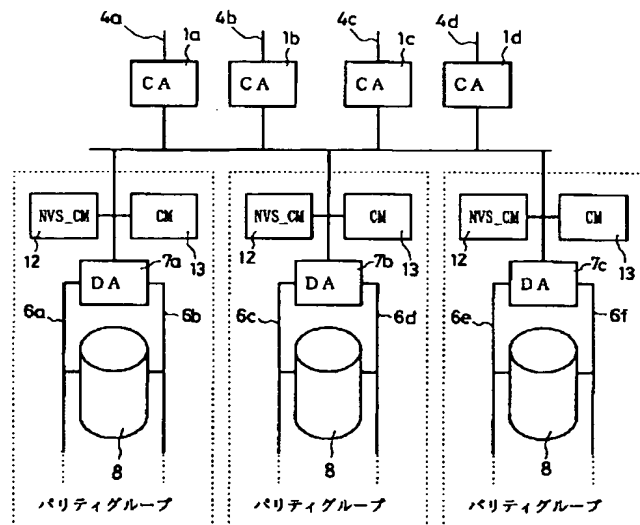
図 7

上位 (CA) 側からのデータ書き込み要求処理



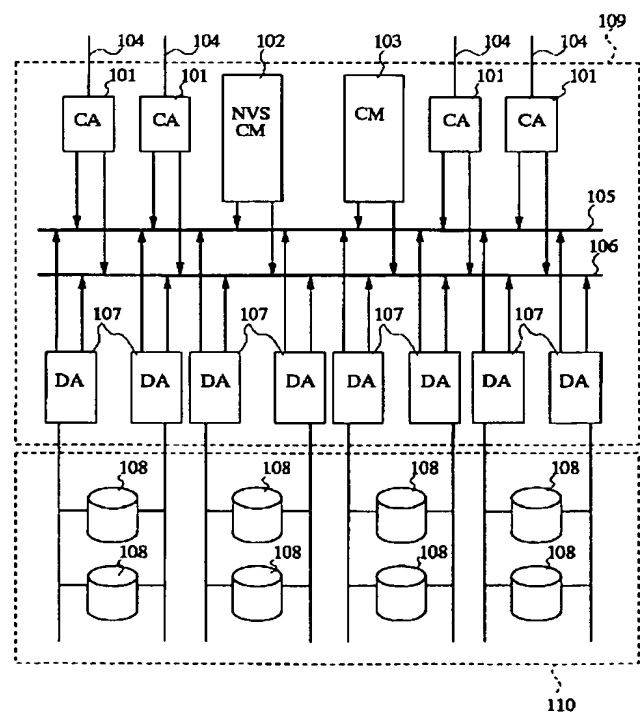
【図 12】

図 12



【図 13】

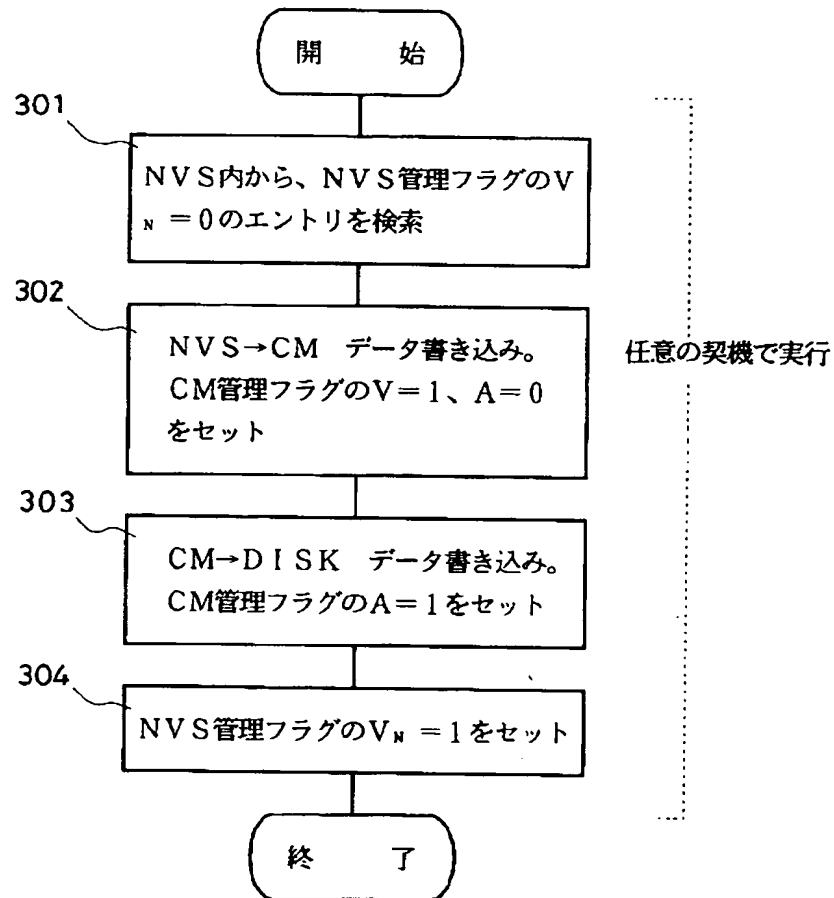
図 13



【図8】

図 8

NVSからCMおよびDISK側への書き込みデータ反映処理



【図9】

図 9

上位(CA)側からのデータ読み出し要求処理

